

Geir-Harald Strand and Vilni Verner Holst Bloch

Statistical grids for Norway

Documentation of national grids for
analysis and visualisation of spatial
data in Norway

Content

Preface	3
Summary	4
1. Introduction	6
1.1. Background.....	6
1.2 Purpose and advantages.....	6
1.3. The role of the SSBgrid in the Norwegian geospatial data infrastructure.....	7
2. Definitions	8
2.1 SSBgrid.....	8
2.2 Projection and datum.....	8
2.3 Truncation, concatenation and rounding.....	9
2.4 Alignment.....	10
2.5 Identification.....	12
2.6 Extention.....	13
2.7 Grid Cell Centre Coordinate.....	14
2.8 Storage and exchange format.....	15
2.9 Size restrictions and identification of local files.....	16
2.10 Basic regional unit and SSBgrid.....	16
3. Spatial models	17
4. Attribute data	19
4.1 Counting.....	19
4.2 Statistics.....	21
4.3 Sample point.....	22
4.4 Area measurement.....	24
4.5 Spatial filters.....	26
4.6 Interpolation.....	27
4.7 Local statistics.....	27
5. Data sources	28
5.1 Population and attributes.....	28
5.2 Buildings and attributes.....	28
5.3 Ground properties and attributes.....	28
5.4 Enterprises and attributes.....	28
5.5 The administrative farms register.....	28
5.6 Norge digitalt.....	29
6. Considerations	30
6.1 Confidentiality.....	30
6.2 Non-equal areas.....	32
6.3 Compatibility with international grids.....	32
6.4 Metadata.....	33
7 References	35
Appendix	36
Statistics by subject and ISO 19115 Topic Categories.....	36

List of figures

Figure 1: Examples of SSBgrids and extensions.....	14
Figure 2: Representation of a SSBgrid as a vector dataset.....	15
Figure 3: Example on use of SSB1KM_F01 for Østfold county	18
Figure 4: Example on SSB5KM_K0420 for Eidskog municipality	19
Figure 5: Barns and cowsheds superimposed on SSB5KM_K0420	20
Figure 6: Example on spatially joined data to SSB5KM_K0420.....	20
Figure 7: Example on aggregated data to SSB5KM_K0420.....	21
Figure 8: SSB5KM_K0420 showing number of buildings (barns and cowsheds).....	21
Figure 9: SSB5KM_K0420 table with buildings (barns and cowsheds). Number and area (m ²)	22
Figure 10: SSB5KM_K0420 showing the mean size in m ² of buildings (barns and cowsheds).....	22
Figure 11: SSB5KM_K0420 with the centre of each grid cell marked.....	23
Figure 12: SSB5KM_K0420 with centre points superimposed on a land resource map	23
Figure 13: SSB5KM_K0420 showing the land resource class at the centre of each grid	24
Figure 14: Part of SSB5KM superimposed on land resource map of Eidskog.	25
Figure 15: Table with land resource types and area measures.	25
Figure 16: SSB5KM_K0420 and agricultural area by grid cell in Eidskog municipality. Per cent.	26
Figure 17. Population by 1 square kilometre grid. Whole of Finnmark.....	30

List of tables

Table 1: Standard grids with national coverage	8
Table 2: ISO 19115 Core Metadata Elements:.....	33
Table 3: ISO 19115 Topic Categories	33
Table 4: Explanation of symbols.....	33

List of boxes

Box 1: Example of determining grid cell in SSB1KM grid	11
Box 2 : Example on finding a grid cell ID in SSB5KM grid.	12

Preface

This document describes and defines the spatial grids used for display and analysis of “raster data” in Statistics Norway. The document is based on previous work by Mona Takle (Takle 2002) and Arne Knut Ottestad (2006 unpublished). Updates, new information on confidentiality rules along with new examples have been made for the present document.

The work with grid-based statistics in Statistics Norway has benefited from previous work in the other Nordic national statistical institutes. Statistics Finland and Statistics Sweden have relatively long experience with grid-based statistics, Finland starting in 1970. Between 1995 and 1997, Statistics Finland and Statistics Sweden ran the first co-operation project on the use of grid data and GIS. This Nordic co-operation entered its next phase in the year 2000 with the first efforts aimed towards constructing a joint map of the population density in the Nordic countries. Later this has been expanded to a pan European initiative (The European Forum for GeoStatistics – www.efgs.ssb.no). By January 2009 the European map comprised Austria, Denmark, Estonia, Finland, Ireland, Norway, Slovenia, Sweden, Switzerland and The Netherlands.

The Norwegian Geospatial Data Infrastructure (NGDI - *Norge digitalt*) was established in 2003-2004. All public institutions producing spatial data, including Statistics Norway, participate in the infrastructure. As the main provider of demographic and socio-economic data in Norway, Statistics Norway has a special responsibility to ensure the compatibility between social science data, data from public registers and the traditional, topographic and physical data provided by other partners in the infrastructure. A statistical standardized grid is an important tool for such compatibility and for the comparability between data sets of quite different contents and structures.

This document has been sent for revision to representatives of institutions participating in the Forum for thematic data (Temadataforum) under NGDI;

- The County Governor in Østfold (FMOS)
- The Directorate for Civil Protection and Emergency Planning (DSB)
- The Directorate for Nature Management (DN)
- The Geological Survey of Norway (NGU)
- The Norwegian Forest and Landscape Institute (Skog og landskap/NFLI)
- The Norwegian Mapping Authority (SK)
- The Norwegian Meteorological Institute (MET)
- The Norwegian Pollution Control Authority (SFT)
- The Norwegian Public Roads Administration (SV)
- The Norwegian Water Resources and Energy Directorate (NVE)

By providing such a national standard, Statistics Norway provide a common reference frame for spatial grid data for use in Norway, and thus promote further integration of data from the different partners in Norwegian Geospatial Data Infrastructure.

Summary

Geo-referenced vector data are point data, linear structures or polygons, and the basic statistical data should be stored in its original form. However, it is often too difficult to analyze and present such data. It is also often necessary to carry out overlay analyses between different data sets to facilitate production of statistics from geo-referenced data. This process creates an even more complex data structure and can also make the thematic maps look totally confusing. A standardized grid structure will facilitate analyses and give a better overview when data are presented on maps. Grids will also be a helpful tool for taking confidentiality aspects into account. For certain statistics and analysis, grid cells can also be used as primary data units, with specific cell identification and statistical information collected for each cell.

SSBgrid is an open-ended definition of a family of spatial tessellation models for use in Norway. The models are all built with quadratic grid cells. The naming convention of the grids is to use the grid cell size, defined as the length of a side of a grid cell, and the unit of measurement (KM for kilometres and M for meters) concatenated to the capital letters 'SSB'. As an example, SSB1KM will be the name of the SSBgrid composed of grid cells with size 1 X 1 kilometer (1 km²). Any quadratic grid size is possible. For grids covering only parts of Norway, a reference to the administrative or other kind of unit can be added to the name.

All SSBgrids are defined in UTM33/WGS84(EUREF89) and aligned in such a way that the southwest corner coordinates $[X_C, Y_C]$ of a grid cell containing position $[X, Y]$ in an SSBgrid with grid cell size K (measured as the length of a side of a grid cell in *meters*) is

$$\left[\text{Truncate}\left(\frac{X + f}{K}\right) \times K - f, \text{Truncate}\left(\frac{Y}{K}\right) \times K \right]$$

where f is a 'false easting' introduced to avoid truncating (see definitions) negative numbers. In the SSBgrid definition, f is set to 2 000 000.

The identification (ID) assigned to each grid cell is the key element of the SSBgrid system. Due to the ID, SSBgrid data can be distributed as tables instead of spatially organized "raster" data. Several datasets using the same SSBgrid can easily be joined together (using the ID as the key) and it facilitates manipulation and analysis of data with standard statistical or tabular data processing tools. The ID of a grid cell with its southwestern corner located at $[X_C, Y_C]$ is;

$$ID = 2 \times 10^{13} + X_C \times 10^7 + Y_C$$

where

X_C is the (UTM33/WGS84(EUREF89)) easting in meters and

Y_C is the (UTM33/WGS84(EUREF89)) northing in meters.

See also the example in box 1.

SSBgrids have no particular exchange format and are most conveniently exchanged using simple tabular formats (text files, EXCEL spreadsheet, SAS or SPSS data files) where each grid cell is represented as a row in the table. A dataset *must* include the ID of each grid cell and *can* contain any number of additional data.

For exchange as SOSI files (the national geodata exchange format), SSBgrid data are handled as tables while SSBgrid vector topologies are handled as polygon data.

Regarding confidentiality, Statistics Norway has so far concluded that; for population statistics (number of persons), exact values can so far not be released for grid cells containing less than 10 persons. Grid cells will therefore operate with the values 0, 1-9, 10, 11 and 12 and so on. For grid cells with 1-9 persons the value is set to 5. Other confidentiality rules can be set for other datasets (see also chapter 6.1).

Grid data on other statistics, like geo-referenced business statistics, income statistics, education statistics, building statistics, land use statistics (e.g. access to shores, noise exposition) and agriculture statistics, must be evaluated by the board of confidentiality.

To get around problems of confidentiality, one might optionally aggregate a number of grid cells to a bigger grid unit consisting of a number of basic cells. Aggregation routines must follow the same principles for confidentiality as for the single grid cells. Aggregated cells should normally be quadratic in order to fit the general SSBgrid model, but exemptions are possible though complicated.

Basis is a square net on 1×1 km cells. For certain areas (cities, intensively used areas), even smaller units, like 500×500 m or 100×100 m, might be an option, when not in conflict with confidentiality rules.

The grids defined in this document are intended for national use in Norway. Other grids will be required for international use. The Protocol on Long-term Financing of the Cooperative Program for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe (EMEP) is using two different grids (one with 50×50 km cells and one with 150×150 km cells) both based on a polar-stereographic projection (www.emep.int). The Joint Research Centre of the European Union and the European Environmental Agency are using a grid based on the Lambert Azimutal Projection (ETRS-LAEA) as described in EEA (2006). In order to deliver national data to these or similar international grids, transformation can be done on demand following a method approved by the receiving agency.

The possibility of aggregation to grids of different kinds, illustrates the need for basic statistical information to be geographically localised as precisely as possible.

1. Introduction

1.1. Background

Administrative and statistical divisions as counties, municipals and *basic regional units* have variable sizes and shapes. They can contain large uninhabited areas as well as areas with high population density. The *basic regional units* are the smallest regional units on which Statistics Norway publish statistics. These units also give the most detailed picture of the settlement pattern.

One of the criteria for delimiting the *basic regional units* was that they should be homogeneous (for example when it comes to population structure); however some of the *basic regional units* contain both rural area and densely built up areas. It can therefore be hard to give a detailed picture of the actual settlement pattern.

Irrespective of whether the objective of the *basic regional units* is fulfilled or not, these geographical units are polygons with irregular size and shape. The geometrical variation is a problem that must be accounted for when data are described or analysed. A uniform grid is more convenient for analysis of spatial patterns or spatial correlation between variables.

Statistics Norway presented population statistics on 250 x 250 meter grids in the municipal reports of The Population and Housing Census 2001. However, the first use of grids in Statistics Norway was in 1875 (Lie and Roll-Hansen 2001, p 284-285)

Today it is possible to geo-reference (geographically locate) people, enterprises, buildings, etc., with their respective attributes. Examples of attributes could be gender, age, education and income for people, number of employees or type of economic activity for enterprises, and size and type of use for buildings.

Connection of personal id-number, id-number for enterprises, id-number for buildings etc. to a numerical address with coordinates, allows a range of socio-economic variables to be spatially referenced and therefore also presented geographically. It is therefore also possible to aggregate statistics to other spatial units than those which traditionally has been used for compilation, distribution and presentation of statistics. One example is to assign statistics to a uniform tessellation model like square grid cells. The advantage of such an approach is that all the units have the same shape and size.

1.2 Purpose and advantages

The purpose of establishing SSBgrid as a national standard is to harmonise the use of grids as units of information at appropriate geographical levels, in order to facilitate integration of data from different sources and meet the needs of different types of users.

SSBgrid is an open-ended definition of a whole family of spatial tessellation models, with quadratic grid cells. As a consequence, everyone can generate and use these grids. SSBgrid supports standardization and data exchange among users and producers of geographical data. SSBgrid thus also facilitates development of new information through combination of data from different sources. Finally, SSBgrid is independent of proprietary formats and can even be used without access to standard GIS software. With SSBgrid, it is easy to use common office software, like Excel and Access, to generate grid statistics and analysis based upon geo-referenced data.

The naming convention of the grids makes it easy to communicate grid cell size and coverage, without having to examine the data in a GIS software package.

The identification (ID) assigned to each grid cell facilitates easy interchange and combination of information from different sources.

SSBgrids, having no particular exchange format, may be easily established in any GIS or statistical system, but are most conveniently exchanged using simple tabular formats.

Grids or other tessellation models will under no circumstances replace other types of statistical divisions such as *basic regional units*. Grids will, however, open up new application areas for traditional statistics and provide a framework for increased collaborative use of statistical data and other spatial data. Statistics linked to grid cells are useful for several reasons.

- it can provide a more detailed spatial resolution compared to other statistical presentation
- the regular shape facilitates analysis (as geographical unit size is a constant, not a variable)
- comparison between different years is simplified as unit size and shape does not change
- different types of data that lack a logical/numerical link can be combined because a geographical link is established through the common grid

1.3. The role of the SSBgrid in the Norwegian geospatial data infrastructure

The Norwegian geospatial data infrastructure (*Norge digitalt*) was established in 2003-2004 following an initiative from the Ministry of Environment (St.mld nr. 30: 2002-2003). All public institutions producing spatial data, including Statistics Norway, are members of the infrastructure.

Statistics Norway is the main provider of demographic and socio-economic data in Norway. Statistics Norway therefore has a special responsibility to ensure the compatibility between social science data and the traditional, topographic and physical data provided by other partners in the infrastructure. A standardized grid is one way to provide such compatibility.

By setting the *SSBgrid* as a standard, Statistics Norway intends to provide a common reference frame for spatial grid data for use in *Norge digitalt* and thus promote further integration of data from the different partners in the spatial infrastructure.

2. Definitions

2.1 SSBgrid

SSBgrid is an open-ended definition of a family of spatial tessellation models for use in Norway. The models are all built with quadratic grid cells. The naming convention of the grids is to use the grid cell size K , defined as the length of a side of a grid cell, and the unit of measurement (KM for kilometres and M for meters) concatenated to the capital letters 'SSB'.

Comments

SSB100M is the SSBgrid composed of grid cells with size 100×100 meter.

For SSB100M, $K = 100$.

Each grid cell in SSB100M is $100^2 = 10\,000$ m²

SSB1KM is the SSBgrid composed of grid cells with size 1×1 kilometer (1000×1000 meter).

For SSB1KM, $K = 1000$.

Each grid cell in SSB1KM is $1000^2 = 1\,000\,000$ m² = 1 km²

Any grid size can be used for an SSBgrid, but the following (table 1 below) are recommended as standard grids with a national coverage:

Table 1: Standard grids with national coverage

SSB100M ⁽¹⁾	Cell size 0.01 km ²	National dataset: Approx. 35 000 000 cells
SSB125M ⁽¹⁾	Cell size 0.015625 km ²	National dataset: Approx. 20 000 000 cells
SSB250M ⁽¹⁾	Cell size 0.0625 km ²	National dataset: Approx. 5 600 000 cells
SSB500M ⁽¹⁾	Cell size 0.25 km ²	National dataset: Approx. 1 400 000 cells
SSB1KM	Cell size 1 km ²	National dataset: Approx. 350 000 cells
SSB5KM	Cell size 25 km ²	National dataset: Approx. 15 000 cells
SSB10KM	Cell size 100 km ²	National dataset: Approx. 15 000 cells
SSB25KM ⁽²⁾	Cell size 625 km ²	National dataset: Approx. 500 cells
SSB50KM ⁽²⁾	Cell size 2 500 km ²	National dataset: Approx. 150 cells
SSB100KM ⁽²⁾	Cell size 10 000 km ²	National dataset: 69 cells
SSB250KM ⁽²⁾	Cell size 62 500 km ²	National dataset: 18 cells
SSB500KM ⁽²⁾	Cell size 250 000 km ²	National dataset: 8 cells

⁽¹⁾ Because of limitations in many software packages and for practical use, these grids are recommended as grids with a county coverage.

⁽²⁾ These grids might also cover sea territories. Number of cells refers to coverage of Norwegian mainland. One has however to be aware of deviations in grid cell areas for regions remote from the Norwegian mainland and Svalbard.

2.2 Projection and datum

All SSBgrids are defined in UTM33/WGS84(EUREF89). The origo of the grid is defined through the use of the UTM coordinate system, zone 33 and the WGS84 datum. UTM33, for short, is the standard way of projecting national geodatasets in Norway.

The Universal Transverse Mercator (UTM) coordinate system is both a set of cartographic projections and a coordinate system for specifying locations on the surface of the Earth. It is used to identify the positions of locations on the earth, but differs from the traditional method of latitude and longitude in several respects. The UTM system is not a single map projection. The system instead employs a series of sixty zones, each of which is based on a specifically defined secant transverse Mercator projection.

The World Geodetic System (WGS) is a standard for use in cartography, geodesy, and navigation. It comprises a standard coordinate frame for the Earth, a standard spheroidal reference surface (the *datum* or *reference ellipsoid*) for raw altitude data, and a gravitational equipotential surface (the *geoid*) that defines the "nominal sea level". WGS84 is the reference system used in the global positioning system NAVSTAR, used by most GPS equipment today.

In 1993 the European Reference Frame 1989 (EUREF89) became the official geodetic datum in Norway. EUREF89 originates by «freezing» WGS84 under an international survey in 1989. A popular explanation of the relationship is that WGS84 is changing continuously due to continental drift and other processes in the physical earth. EUREF89 is an instance of WGS84 as it was at one certain point in time. For maps in the scale 1:50 000 EUREF89 is almost identical with WGS84. In the map series N50 the term EUREF89 (WGS84) is used for geodetic datum.

SSBgrid is using one of the (sixty different) transverse Mercator projections available in the UTM system. The chosen projection is UTM33, because this is the projection commonly used by the Norwegian Mapping Authority (Statens kartverk) to produce maps with a national coverage. The central meridian of UTM33 is the 15° east meridian.

The “origo” of the SSBgrid is hence where the 15° east meridian is crossing the equator.

Comments

Grids defined in UTM33/WGS84(EUREF89) can be *drawn* in other projections, but the grid cells will then not appear as quadratic squares. The “origo” of the grid is the point where the 15° east meridian is crossing the equator. The “x” (east) coordinate for this point is, however [500000,0] (and not [0,0]) because a false easting of 500 000 is added by the definition of the UTM system.

For calculation of IDs, SSBgrid also adds another false easting of 2 000 000 in order to avoid using negative numbers in the identification of the grid cells. The ID of an SSBgrid gridcell with its southwest corner at the “origo” is therefore 25 000 000 000 000 (irrespective of cell size). Such a grid cell would, however, never be found in an actual dataset since the location is far from any Norwegian territory.

2.3 Truncation, concatenation and rounding

In mathematics, *truncation* is the term for limiting the number of digits right of the decimal point, by discarding the least significant ones. For example, consider the real numbers

5.6341432543653654
32.438191288
6.3444444444444

To *truncate* these numbers to 4 decimal digits, we only consider the 4 digits to the right of the decimal point. The result would be:

5.6341
32.4381
6.3444

Note that in some cases, truncating would yield the same result as rounding, but truncation does not round up or round down the digits; it merely cuts off at the specified digit. The truncation error can be twice the maximum error in rounding.

In SSBgrid, truncation is used to determine the westernmost coordinate of a grid cell when the cell size is known. This method works well as long as the coordinates are positive numbers. It does not work when the coordinates are negative numbers. In the projection system UTM33/EUREF89, a significant part of western Norway, as well as the ocean areas west of the Norwegian coast, is described using negative numbers as coordinates. It is therefore necessary to add a positive number (known as a “false easting”) to the coordinates in order to ensure that all coordinates are positive numbers.

“False easting” is a well known and often employed method in cartographic projections. The UTM system always adds a “false easting” of 500 000 to its coordinates. SSBgrid merely adds another 2 000 000 to this “false easting”, ensuring that the employment of truncation works consistently throughout the system.

In computer programming, *string concatenation* is the operation of joining two character strings end to end. For example, the strings "foo" and "bar" may be concatenated to give "foobar".

2.4 Alignment

The southwest corner coordinates $[X_C, Y_C]$ of a grid cell containing position $[X, Y]$ in an SSBgrid with grid cell size K is :

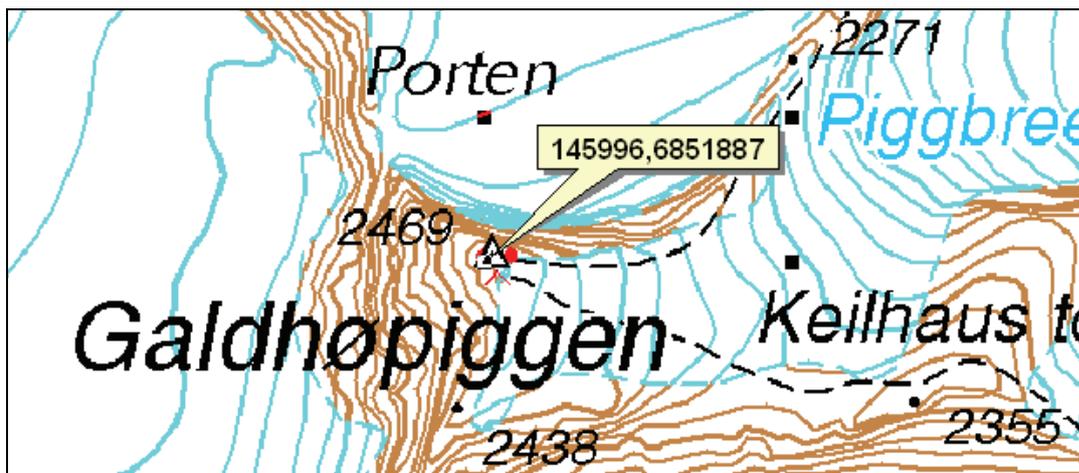
$$\left[\left(\text{Truncate} \left(\frac{X + f}{K} \right) \right) \times K - f, \text{Truncate} \left(\frac{Y}{K} \right) \times K \right]$$

,where f is a ‘false easting’ introduced to avoid truncating negative numbers. The false easting is set to 2 000 000. The false easting is set so that it should not be confused with common coordinates.

Comments

There is (theoretically) an infinite number of possible SSBgrids, but only one possible realization for each given grid cell size.

Box 1: Example of determining grid cell in SSB1KM grid



Base map: N50 raster © Norge digitalt

The coordinates of the summit of Galdhøpiggen mountain is [145996,6851887] (when measured as meter coordinates in UTM33/WGS84(EUREF89)). In order to find the southwest corner of the SSB1KM grid cell (grids with cell size of 1 km²) containing the summit, use the following procedure:

For the X (east) coordinate:

$$\begin{aligned} X+f &= 145\,996 + 2\,000\,000 = 2\,145\,996 \\ (X+f)/K &= 2\,145\,996/1\,000 = 2\,145.996 \\ \text{Truncate}((X+f)/K) &= \text{Truncate}(2\,145.996) = 2\,145 \\ \text{Truncate}((X+f)/K) \times K &= 2\,145 \times 1\,000 = 2\,145\,000 \\ \text{Truncate}((X+f)/K) \times K - f &= 2\,145\,000 - 2\,000\,000 = 145\,000 \end{aligned}$$

For the Y (north) coordinate:

$$\begin{aligned} Y/K &= 6\,851\,887/1\,000 = 6\,851.887 \\ \text{Truncate}(Y/K) &= \text{Truncate}(6\,851.887) = 6\,851 \\ \text{Truncate}(Y/K) \times K &= 6\,851 \times 1\,000 = 6\,851\,000 \end{aligned}$$

The coordinates of the southwest corner of the SSB1KM grid cell containing the summit is thus [145000,6851000]] (when measured as meter coordinates in UTM33/WGS84(EUREF89)).

The ID of this grid cell is furthermore

$$2 \times 10^{13} + X_{\text{CORNER}} \times 10^7 + Y_{\text{CORNER}}$$

To calculate the ID when the southwest corner is known use the following procedure

2×10^{13}	=	20 000 000 000 000
$X_{\text{CORNER}} \times 10^7 = 145\,000 \times 10^7$	=	1 450 000 000 000
$\underline{Y_{\text{CORNER}}}$	=	6 851 000
SUM	=	21 450 006 851 000

$$\text{ID} = 21450006851000$$

Note: The 'false easting' (*f*) is required in order to (a) ensure correct "truncation" of negative coordinates and (b) avoid having to use minus sign in the ID. For SSBgrid, 2 000 000 is the chosen value for this parameter.

2.5 Identification

The identification (ID) of a grid cell with its southwestern corner at $[X_C, Y_C]$ in an SSBgrid with grid cell size K is

$$ID = 2 \times 10^{13} + X_C \times 10^7 + Y_C$$

Comments

Every grid cell in an SSBgrid has a unique, 14 digit ID. The first seven digits is the east coordinate of the southwestern (lower left) corner of the grid cell (as meter coordinate in UTM33/WGS84(EUREF89)) *with an additional false easting of 2 000 000*. The false easting is added in order to avoid problems generated by negative east coordinates and is added to the standard false easting of the UTM coordinate. The last seven digits is the north coordinate of the southwestern (lower left) corner of the grid cell (as meter coordinate in UTM33/WGS84(EUREF89)).

The ID is the key element of the SSBgrid system. Due to the ID, SSBgrid data can be distributed as tables instead of spatially organized “raster” data. Several datasets using the same SSBgrid can easily be joined together (using the ID as the key) and it facilitates manipulation and analysis of data with standard statistical or tabular data processing tools (e.g EXCEL or SPSS).

Computational efficiency

The ID can be calculated directly from point coordinates for any location within the grid cell. The false easting, needed to compute X_C and Y_C for locations along the western coast of Norway, must also be employed when the ID is calculated directly:

$$ID = Trunc \left[\frac{X + 2 \times 10^6}{K} \right] \times K \times 10^7 + Trunc \left[\frac{Y}{K} \right] \times K$$

Box 2 : Example on finding a grid cell ID in SSB5KM grid.

The coordinates of the summit of Galdhøpiggen mountain is $[145996, 6851887]$ (when measured as meter coordinates in UTM33/WGS84(EUREF89)) (see also map in Box 1). The ID for the grid cell containing this location in SSB5KM is

$$\begin{aligned} ID &= Trunc \left[\frac{(X+2 \times 10^6)}{K} \right] \times K \times 10^7 + Trunc \left[\frac{Y}{K} \right] \times K \\ &= Trunc \left[\frac{(145996+2 \times 10^6)}{5000} \right] \times 5000 \times 10^7 + Trunc \left[\frac{6851887}{5000} \right] \times 5000 \\ &= Trunc \left[\frac{2145996}{5000} \right] \times 5000 \times 10^7 + Trunc \left[\frac{6851887}{5000} \right] \times 5000 \\ &= Trunc \left[429.1992 \right] \times 5000 \times 10^7 + Trunc \left[1370.3774 \right] \times 5000 \\ &= 429 \times 5000 \times 10^7 + 1370 \times 5000 \\ &= 2145000 \times 10^7 + 6850000 \\ &= 21450006850000 \end{aligned}$$

$$ID = 21450006850000$$

Notice that this ID – calculated for SSB5KM - is different from the ID calculated for the same location in SSB1KM in the example in Box 1 above.

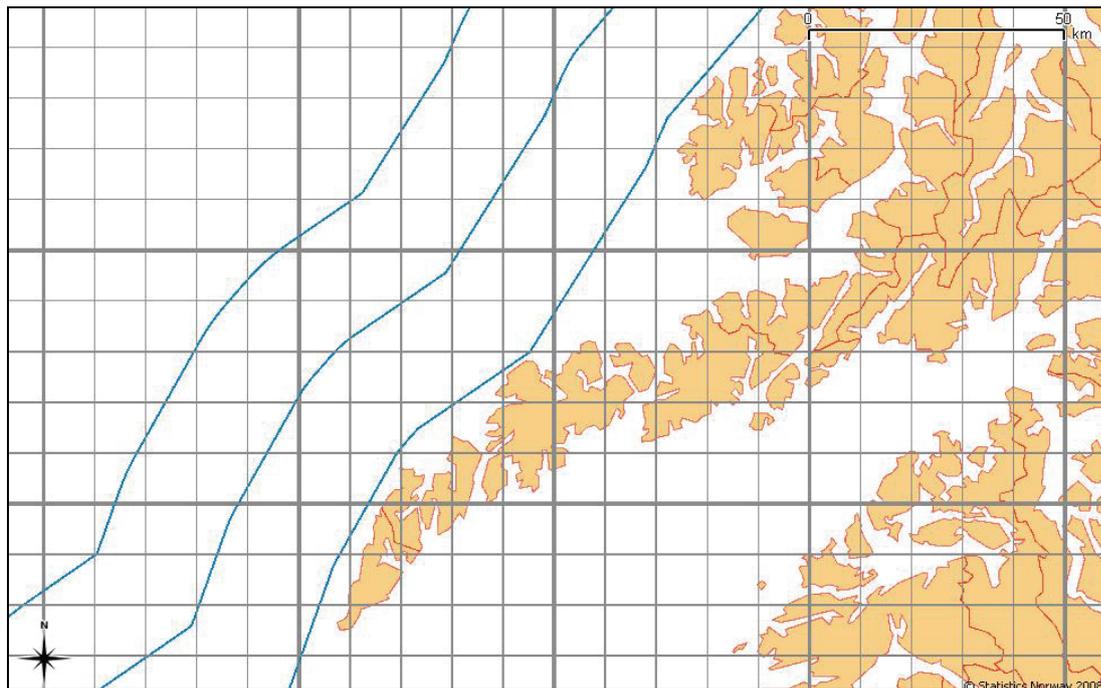
2.6 Extention

SSBgrids have no definite extension.

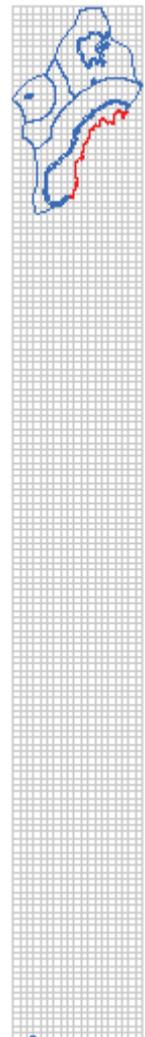
Comments

To cover the Norwegian Mainland, it is recommended to start at a point southwest of the location represented as [-80 000,6 440 000] (to include the southernmost and westernmost points of the Mainland Norway) and include an area reaching at least 1 200 000 meters eastward and 1 500 000 meters northward. To include Svalbard, it is recommended that the area is further extended 1 100 000 meters northward (Figure 2).

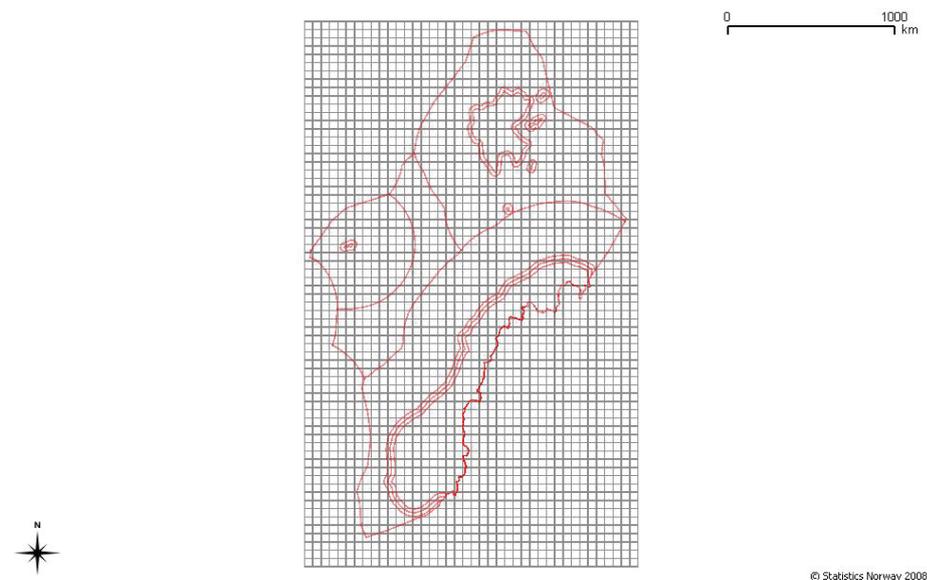
Figure 1: Examples of SSBgrids and extensions.



(a) Example: SSB10KM (light grey lines) and SSB50KM (dark grey lines) grid in Lofoten



(c) SSB100KM grid covering Norwegian territories in the northern and southern hemisphere (Bouvet Island in Antarctica being the blue dot in the bottom left)



(b) SSB50KM grid covering Norwegian sea territories in the northern hemisphere

2.7 Grid Cell Centre Coordinate

The centre coordinates of an SSBgrid cell with identification ID and cell size K is

$$X_C = \text{Truncate}(ID \times 10^{-7}) - 2 \times 10^6 + \frac{K}{2}$$

$$Y_c = ID - \text{Truncate}(ID \times 10^{-7}) \times 10^7 + \frac{K}{2}$$

Comments

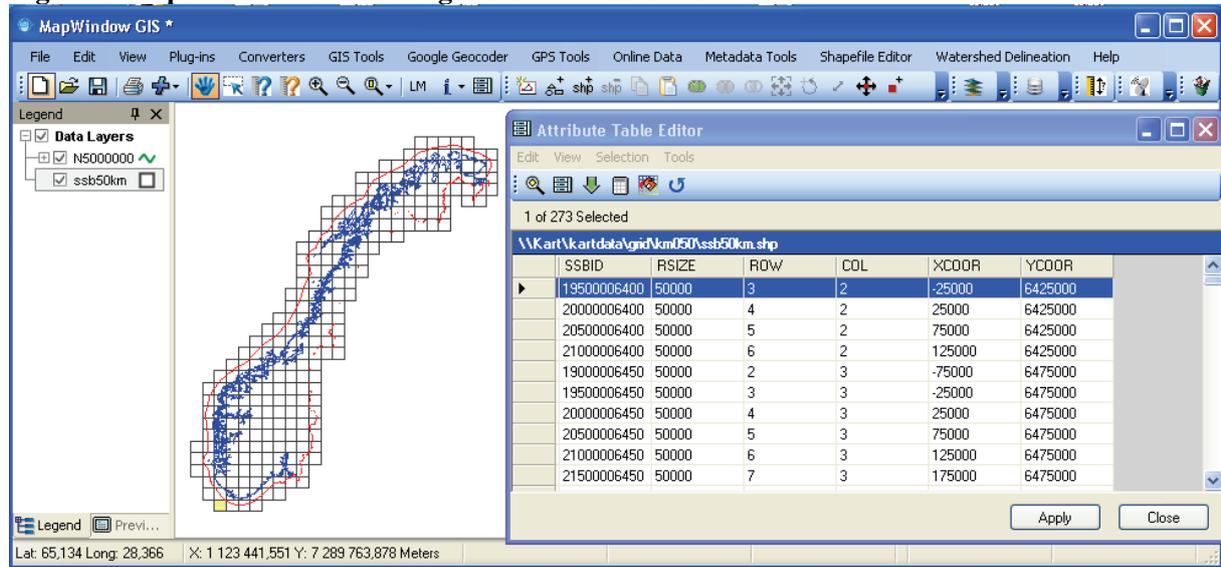
Other points in a grid cell can be calculated accordingly.

2.8 Storage and exchange format

SSBgrids have no particular storage or exchange format. A further advantage of the SSBgrid approach is that only a single geospatial dataset is required for each grid size. The geospatial dataset can be established as a vector data model with each grid cell represented as a (simple, square) polygon labelled with the appropriate ID but no further data attached. For visualisation or other GIS operations, the table with thematic data should be merged with the geospatial dataset using a database *join* operation with ID as the foreign key.

The representation of an SSBgrid as a vector dataset is illustrated in Figure 3. Each grid cell is represented as a square polygon (shown in red in the left frame – an ordinary map in grey is added for reference). The same dataset can be displayed as a table (shown in the right frame). When a grid cell is selected (shown as a small yellow square in the left frame) it is simultaneously selected in the tabular appearance of the data set (shown as a yellow row in the right frame). Additional data (from a spreadsheet or any other tabular structure) can be added to (technically: joined into) the tabular structure in the right frame and used for visualization as thematic maps (in the left frame).

Figure 2: Representation of a SSBgrid as a vector dataset.



Note: SSB50KM as a vector data set with each grid cell represented as a square polygon. In a GIS, this dataset can be displayed both as a map (left) and a table (right).

Comments

SSBgrid data are most conveniently exchanged using simple tabular formats (text files, EXCEL spreadsheet, SAS or SPSS data files) where each grid cell is represented as a row in the table. A dataset *must* include the ID of each grid cell and *can* contain any number of additional data.

For exchange as SOSI files (the Norwegian standard exchange format), SSBgrid data are handled as tables.

2.9 Size restrictions and identification of local files

The capacity of Excel 2007 is limited to 1 048 576 rows by 16 384 columns. SSB1KM will, for the Norwegian Mainland, require approximately 350 000 rows while SSB500M will require more than 1.3 million rows. It is therefore possible to store national grid data down to a cell size of 1 km² as Excel files. More detailed data require other formats, but are not likely to be used much for national datasets – partly because the dataset will be too large, partly because such small grids do not communicate on a national level, and partly because small grid cells generate conflicts with confidentiality. For more detailed grids the following standard naming convention and coverage is to be used:

SSB100M_F01 ... SSB100M_F21 ...
SSB250M_F01 ... SSB250M_F21 ...
SSB500M_F01 ... SSB500M_F21 ...

and

SSB100M_K0101 ... SSB100M_K2101 ...
SSB250M_K0101 ... SSB500M_K2101 ...
SSB250M_K0101 ... SSB500M_K2101 ...

Where ‘F’ and ‘K’ in the prefix stands for ‘Fylke’ (county) and ‘Kommune’ (municipality) respectively, followed by a code for corresponding county or municipality.

The advantage of using Excel or similar, tabular structures for storage and distribution of data is that a large number of thematic data can be conveniently stored in a single data set. Each row in the table represents one grid cell, the columns represents the different thematic attributes. The ID is stored in a separate column. Analysis can be carried out in the spreadsheet, or the table can be imported into a statistical package (SPSS, SAS, Minitab etc).

2.10 Basic regional unit and SSBgrid

Basic regional units are the smallest regional units on which Statistics Norway publish statistics. These units are delineated using natural and societal borders; i.e. rivers, coastlines, administrative borders and ground property borders.

Comments

One of the criteria for delimiting the basic regional units was that they should be homogeneous (for example when it comes to population structure); however some of the basic regional units contain both rural area and densely built up areas. It can therefore be hard to give a detailed picture of the actual settlement pattern. The SSBgrid offers an alternative structure for geographical presentation of statistics for small areas.

3. Spatial models

The spatial representation of the SSBgrids can be

- Point data sets (representing the centre of each grid cell)
- Polygon data sets (representing each grid cell as a polygon)
- Raster data sets

Polygon data sets are the main spatial representation of SSBgrids used by Statistics Norway. In a polygon model of a grid, each individual grid cell is represented as a separate polygon. Neighbouring grid cells must share polygon boundaries (i.e. no gap or overlap is allowed). The polygon model can be built and distributed independent of the attribute data as long as the unique and standardized ID is attached to each grid cell polygon. The spatial model of a grid with a specific resolution can therefore be generated once and reused by any number of users. It is not necessary to generate and distribute the polygon model together with the attribute data.

Polygon representations of the most common SSBgrids are produced by Statistics Norway in SOSI and ESRI Shape formats. These data can be downloaded from www.ssb.no and will also be available through the distribution service operated by the Norwegian geospatial data infrastructure *Norge digitalt*.

Point data sets are also convenient models for grid cartography. While polygon models only allows variation in the colour and texture of each grid cell, point models also support the use of variable symbol, shape and size for drawing grids. These principles may be combined for analytical purposes. An example is the map of population pressure on agricultural areas (Strand and Rogstad 2004) where square symbols are scaled according to the proportion of agricultural area in the grid cell and a colour assigned to represent the population density (Figure 5), putting pressure on the agricultural areas.

A point data set is easily generated from an SSBgrid since the centre coordinates of each grid cell can be computed using the ID. The method is described in chapter 2.7.

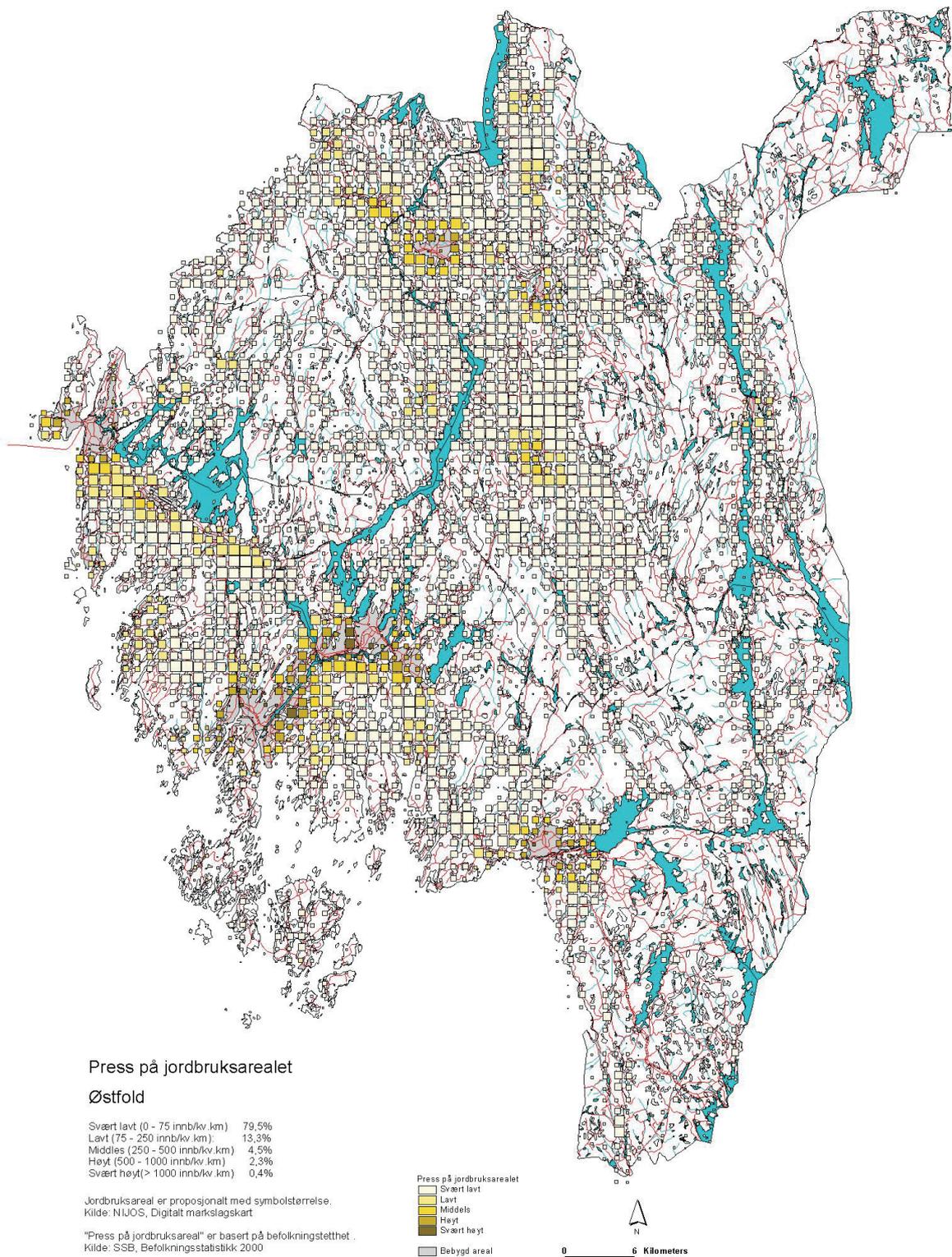
Figure 3 (next page) shows SSB1KM for Østfold county used to analyse and display population density (shown as colour – darker colour represents higher density) and agricultural land (the square symbol drawn in each SSBgrid cell is scaled proportional to the amount of agricultural land in the grid cell).

Raster data set is a possible but inconvenient way to represent SSBgrids. A raster data set is a set of attribute values for a grid, organized in such a way that the position of each attribute value within the data set reflects its location in the spatial model. Information distributed along with the attribute data will explain how to transform positions within the dataset to actual geographical locations. Two problems regarding raster representation of grids are

- (a) it is hard to maintain the ID of the grid cells as they have to be included as an attribute and most raster formats does not support character strings or fourteen digit numbers or more as attribute values.
- (b) raster formats are supported by software that is efficient for displaying grids as images and performing image analytical operations, but inefficient for further statistical analysis and modelling.

A large number of different raster formats exists but are not recommended for general storage and distribution of grid data. Instead, it is recommended to store grid data as tables and to convert these tables to raster data in situations where raster data are required. A convenient method is first to convert the table to a point data set (by using the ID to generate centre coordinates for each grid cell) and then to convert the point data set to a raster using GIS or image processing software.

Figure 3: Example on use of SSB1KM_F01 for Østfold county



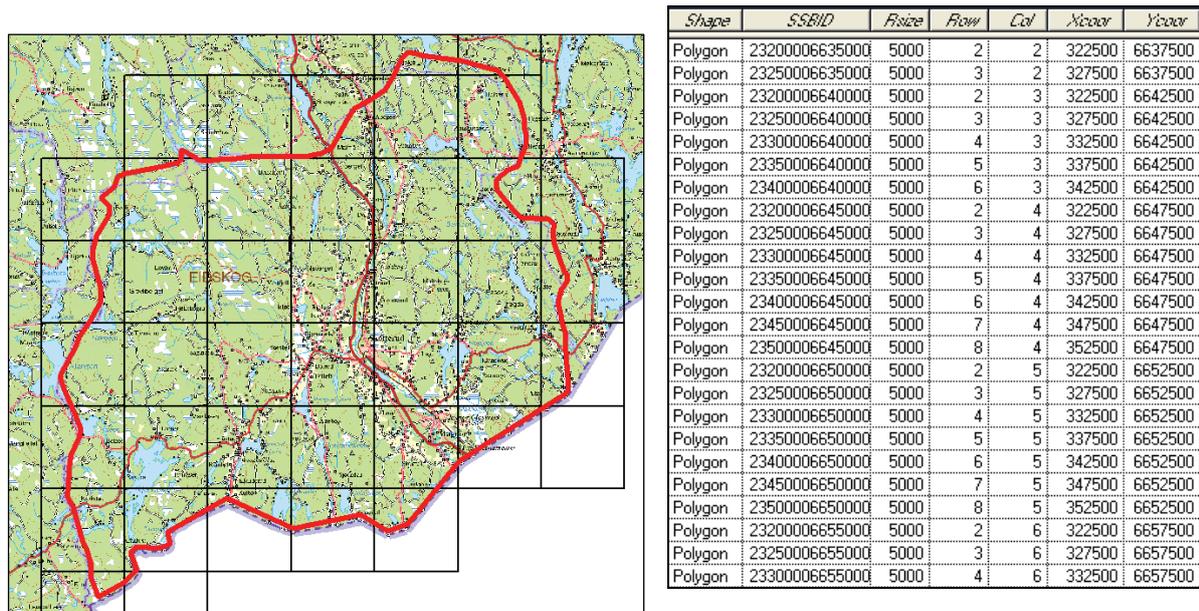
Source: Strand and Rogstad (2004). Base map: N250 © Norge digitalt

4. Attribute data

Data can be assigned to grid cells in numerous ways. The choice of method partly depends on the data source, partly on the purpose of the model. A few, common methods are discussed below, but the list is not exhaustive.

The discussion is illustrated using examples with SSB5KM for Eidskog municipality in southeastern Norway (Figure 4). SSB5KM is in this case a coarse grid relative to the size of the study area, and the example is intended for illustration only.

Figure 4: Example on SSB5KM_K0420 for Eidskog municipality



Eidskog municipality with SSB5KM superimposed (left) and part of SSB5KM for the same area represented as a table (right). Base map: N250 Raster © Norge digitalt.

4.1 Counting

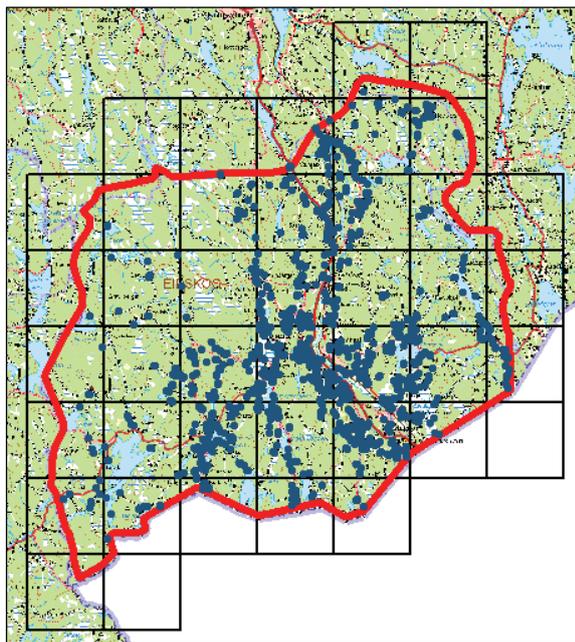
Counting is the method commonly used by Statistics Norway in order to assign attribute data to grids. Each resident in Norway is linked to an address that is furthermore attached to a coordinate. The central population register can thus be interpreted as a large point data set. These point data can be assigned to SSBgrid cells using the calculation of ID explained in box 1. The SSBgrid ID thus becomes an attribute of each person in the population register. Performing a tabular AGGREGATE function on the population register using ID as the BREAK variable and simply COUNTing the number of cases for each break group produces an SSBgrid with population as attribute data.

Clearly, the method can be used with any point data set where the points have coordinates and represent countable cases (objects or events). Examples are car accidents, robberies or buildings.

Counting can also be performed on a *selection* of the original dataset, e.g buildings used as vacation homes.

Figure 5 shows barns and cowsheds in Eidskog. Each building is represented as a point (event). A *spatial join* between the dataset containing the buildings and the SSB5KM dataset will add the ID of the corresponding grid cell to each building. The buildings database is now a table with the grid ID attached, as shown in figure 6.

Figure 5: Barns and cowsheds superimposed on SSB5KM_K0420



Shape	Bygntype	Kommnr	Areal_f
Point	242	0420	73.54495
Point	242	0420	103.83165
Point	241	0420	221.18440
Point	242	0420	65.73910
Point	242	0420	74.63260
Point	242	0420	100.92460
Point	241	0420	197.65775
Point	242	0420	168.88190
Point	241	0420	235.95190
Point	242	0420	64.38030
Point	242	0420	336.84665
Point	242	0420	109.00000

Note: Barns and cowsheds in Eidskog municipality superimposed on SSB5KM (left). The buildings represented as a table (excerpt right). Base map: N250 Raster © Norge digitalt

Figure 6: Example on spatially joined data to SSB5KM_K0420

Shape	Bygntype	Kommnr	Areal_f	SSBID	Rsize	Flow	Col	Xcoord	Ycoord
Point	242	0420	73.54495	23350006645000	5000	5	4	337500	6647500
Point	242	0420	103.83165	23400006650000	5000	6	5	342500	6652500
Point	241	0420	221.18440	23450006650000	5000	7	5	347500	6652500
Point	242	0420	65.73910	23350006665000	5000	5	8	337500	6667500
Point	242	0420	74.63260	23400006650000	5000	6	5	342500	6652500
Point	242	0420	100.92460	23400006645000	5000	6	4	342500	6647500
Point	241	0420	197.65775	23350006660000	5000	5	7	337500	6662500
Point	242	0420	168.88190	23350006660000	5000	5	7	337500	6662500
Point	241	0420	235.95190	23400006650000	5000	6	5	342500	6652500
Point	242	0420	64.38030	23400006650000	5000	6	5	342500	6652500
Point	242	0420	336.84665	23400006645000	5000	6	4	342500	6647500
Point	242	0420	109.00000	23300006640000	5000	4	3	332500	6642500

Buildings (from figure 5) represented as a table (excerpt) after a spatial join with SSB5KM. The corresponding ID (called SSBID in the table) is attached to each building.

The table shown in figure 6 can be aggregated using SSBID as a *break variable* and simply counting the number of cases in each break group. The result is a new table, shown in figure 7 (excerpt).

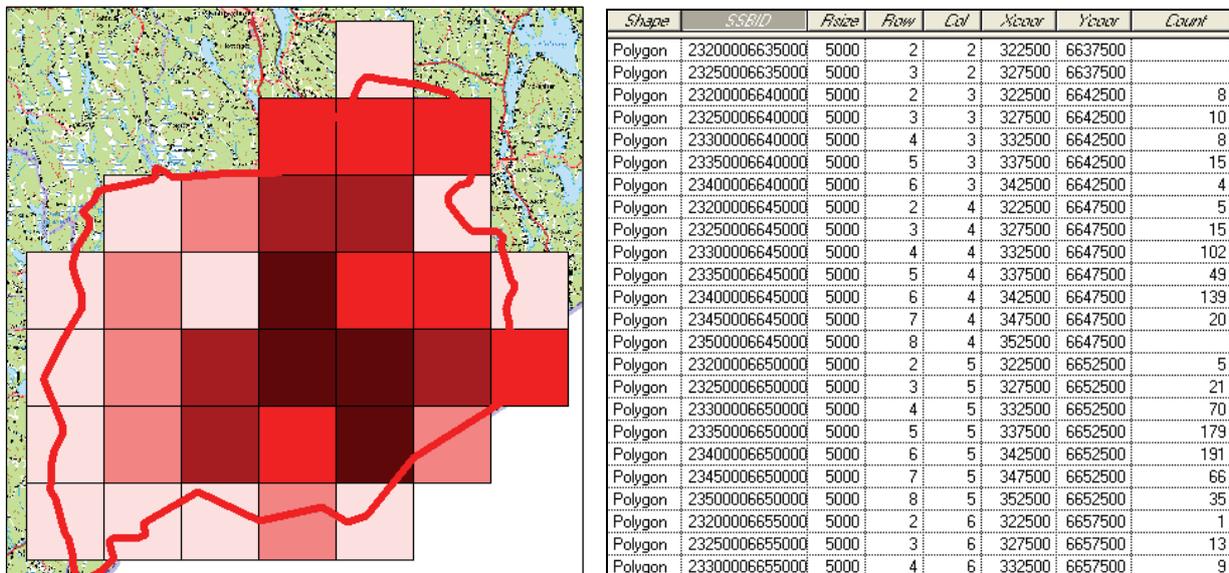
Figure 7: Example on aggregated data to SSB5KM_K0420

SSBID	Count
23200006640000	8
23200006645000	5
23200006650000	5
23200006655000	1
23250006640000	10
23250006645000	15
23250006650000	21
23250006655000	13

Note: Table showing the number of buildings (barns and cowsheds) found in each grid cell of SSB5KM in Eidskog.

The table shown in figure 7 can subsequently be joined with the original SSB5KM table (from figure 4) using SSBID as the common key. The result is that the number of building (in this case limited to barns and cowsheds) becomes an attribute of the SSB5KM dataset and can be used for further analysis and visualization (figure 10).

Figure 8: SSB5KM_K0420 showing number of buildings (barns and cowsheds)



Note: SSB5KM showing the number of buildings (barns and cowsheds) found in each grid cell of SSB5KM in Eidskog. Data displayed as a map (left – darker colour represents more buildings) and a table (right). Base map: N250 Raster © Norge digitalt

4.2 Statistics

The counting strategy is easily extended to include various statistical calculations on the attributes of the cases being counted. The SSBgrid ID is still calculated and assigned an additional attribute of the original point data set, and then used as a break variable to produce a new aggregated dataset. But instead of, or in addition to, counting cases, one or more descriptive statistical functions are carried out on an attribute in the original dataset. Examples are calculating mean income or variance in age in each break group. The result is a new dataset with SSBgrid IDs as cases and the computed statistics as variables.

In the example used to illustrate counting in chapter 4.1, *statistics* can be introduced at the step between figure 6 and 8. When the table in figure 8 is aggregated using SSBID as a *break variable*, the counting function can be replaced by any other available arithmetic or statistical function involving

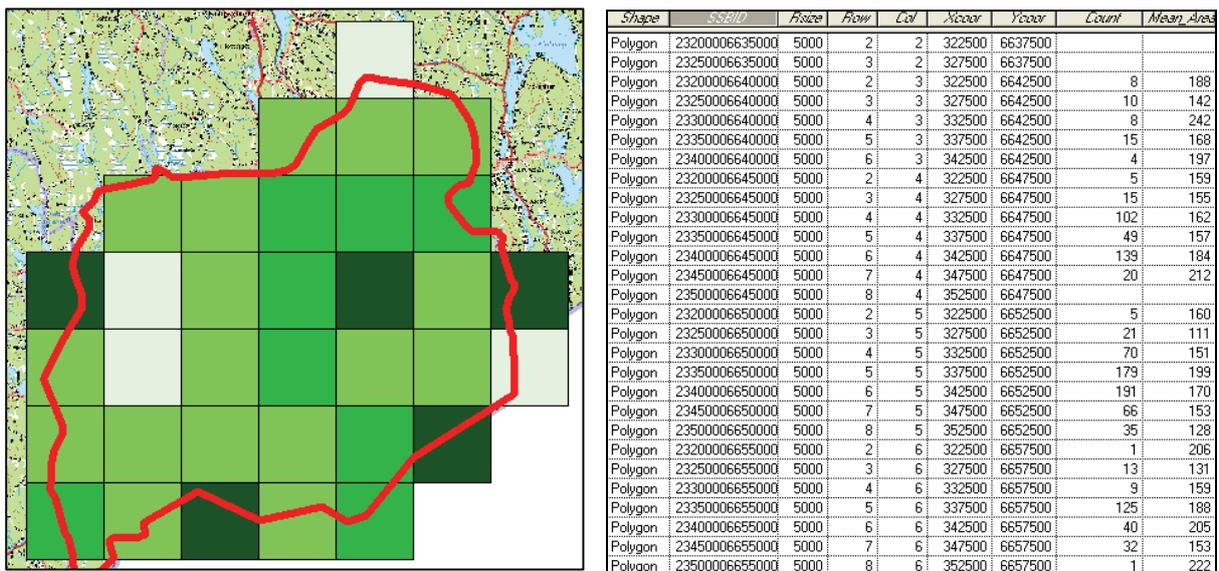
data available in the table in figure 9. An example is to calculate the mean floor area (column Areal_f in figure 6) for each break group. The result is shown in figure 9. It is a table structurally similar to the table in figure 7 but with the extra column added of SSB5KM in Eidskog.

Figure 9: SSB5KM_K0420 table with buildings (barns and cowsheds). Number and area (m²)

<i>SSBID</i>	<i>Count</i>	<i>Mean Area</i>
23200006640000	8	188
23200006645000	5	159
23200006650000	5	160
23200006655000	1	206
23250006640000	10	142
23250006645000	15	155
23250006650000	21	111
23250006655000	13	131

This table can, as in the example above, be joined with SSB5KM for further analysis and thematic map production (figure 10).

Figure 10: SSB5KM_K0420 showing the mean size in m² of buildings (barns and cowsheds)

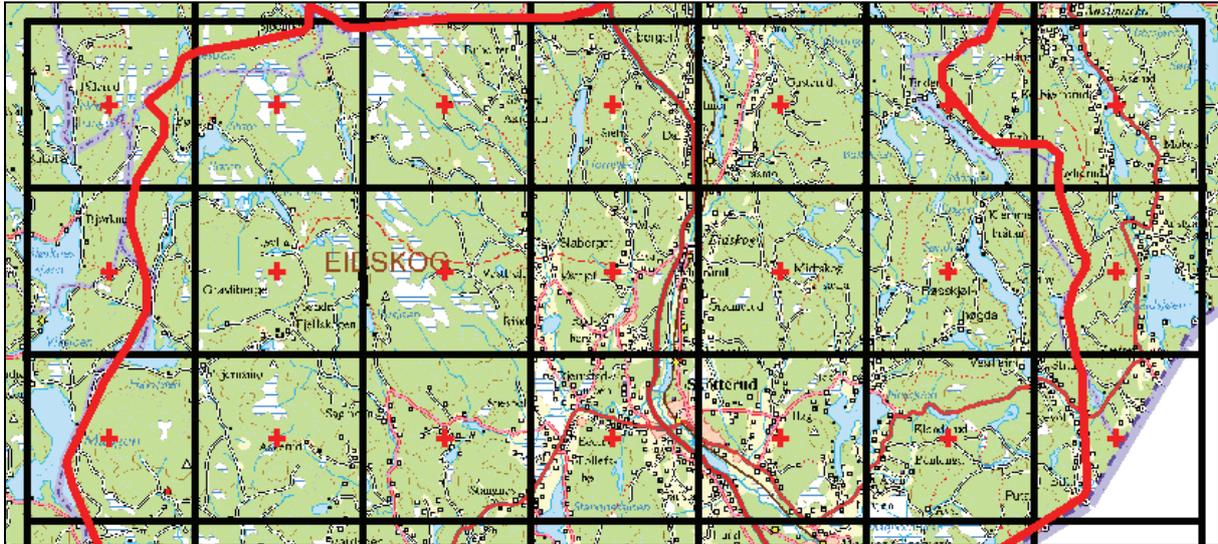


Note: SSB5KM showing the mean size in m² of buildings (barns and cowsheds) found in each grid cell of SSB5KM in Eidskog. Data displayed as a map (left – darker colour represents larger average) and a table (right). Base map: N250 Raster © Norge digitalt

4.3 Sample point

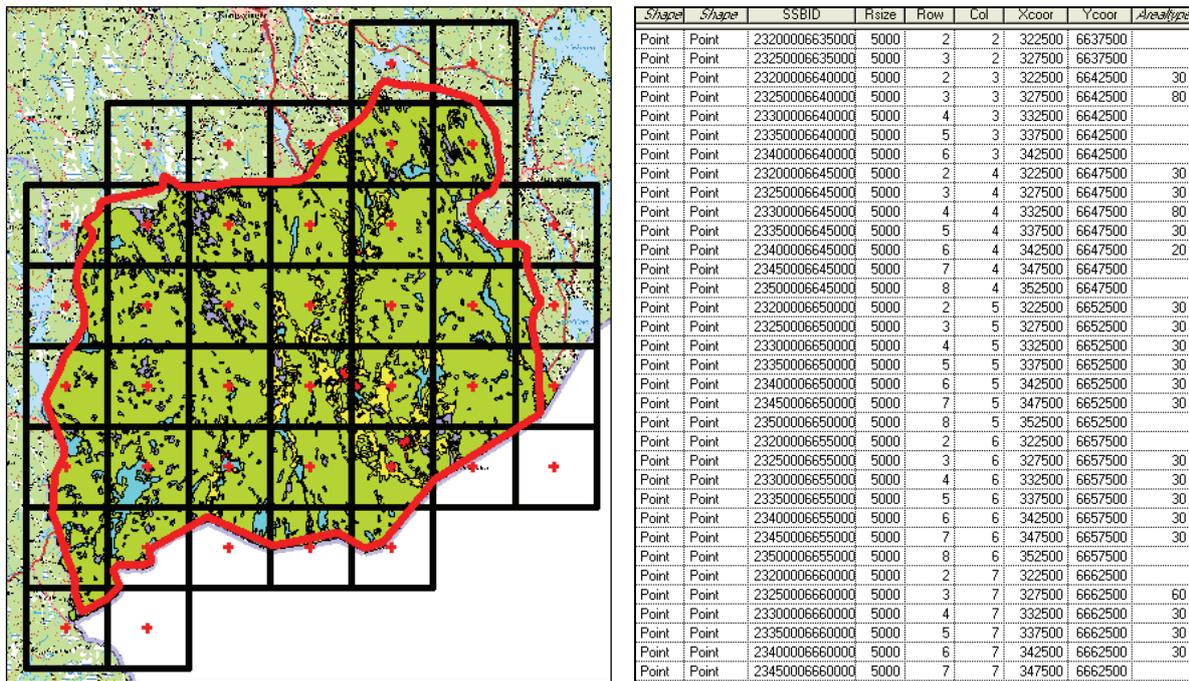
The sample point strategy is used to assign data from an existing, discrete polygon map to a grid. An example is to find the soil type of each grid cell in an SSBgrid. The method is usually employed with the centre point of the grid cell as the sampling point (figure 11), but any point within the grid cell could be used – including the lower left corner or a random point within the grid cell.

Figure 11: SSB5KM_K0420 with the centre of each grid cell marked



Note: Part of SSB5KM in Eidskog with the centre of each grid cell marked. Thematic data (land cover, soil, geology etc) found at the centre can be attached as an attribute of the grid cell. Base map: N250 Raster © Norge digitalt.

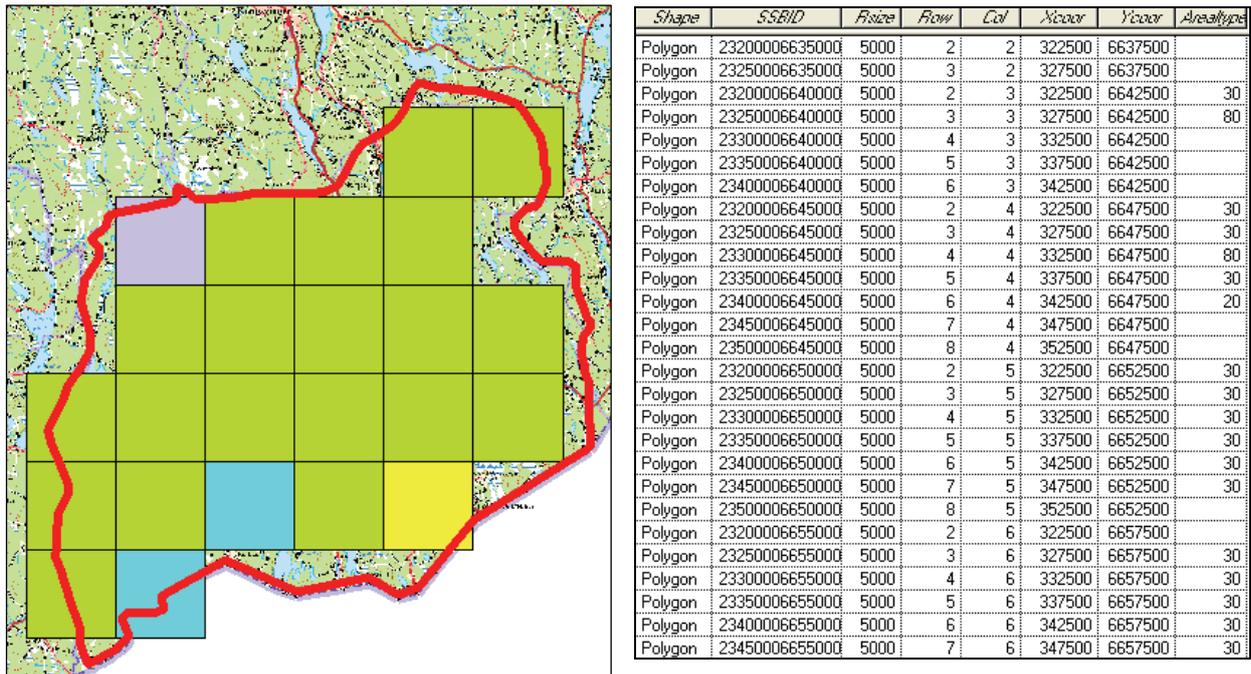
Figure 12: SSB5KM_K0420 with centre points superimposed on a land resource map



Note: SSB5KM with centre points superimposed on a land resource map of Eidskog. Data displayed as a map (left) and a table (right). The land resource class (*Arealtype*) at each point is attached to the table through a *spatial join* operation.

From an SSBgrid, the corresponding set of point coordinates can easily be computed using the methods described in section 4.3 above. This point data set can then be intersected with the polygon map (e.g. a soil map) in order to transfer the attributes of the polygon map (e.g. soil type) to the point data set.

Figure 13: SSB5KM_K0420 showing the land resource class at the centre of each grid



Note: SSB5KM showing the land resource class at the centre of each grid cell in Eidskog. Data displayed as a map (left: Agriculture (yellow), Forest (green), Bogs (Violet) and Water (Blue)) and a table (right). Notice how some grid cells are missing because the centre is located outside the municipal boundary. The grid cells are in this case much too large to convey a meaningful map and the example is only included to illustrate the method. For details on statistics based upon sampling techniques see (Strand and Rekdal, 2006). Base map: N250 Raster © Norge digitalt

The operation is also known as a *spatial join*. In the examples in chapters 4.1 and 4.2 above, a spatial join was used to find the grid cell ID for each building in a database of buildings where the buildings were represented as points. It is the same operation that now is used to find the thematic value from a thematic map and attach it to the centre (point) of each grid cell.

The result of the spatial join is a set of centre points with IDs and thematic values (shown as a table in figure 12). This dataset can be joined to the grid - using the ID as a *foreign key* - for further analysis and visualization (figure 14).

The sample point method is convenient because it is fast and easy to implement and does not require much computing power. A number of attributes linked to the polygons can be transferred to the point data set in a single operation. The limitations of this method is that a grid cell can extend over several different polygons and the method only assigns data from the polygon that happens to intersect with the point coordinate chosen to represent the grid cell. This problem becomes particularly serious when large grid cells are used (e.g. soil type or land cover assigned to SSB1KM or larger, as shown in figure 13).

4.4 Area measurement

The area measurement strategy is an alternative to the sample point method described in section 4.3 above. It is used to assign data from an existing, discrete polygon map to a grid. The method does, however, measure the complete area extent of each polygon class within each grid cell. It is therefore a more accurate method than the sample point method, but is more complex to implement and requires more computing power.

Figure 14: Part of SSB5KM superimposed on land resource map of Eidskog.

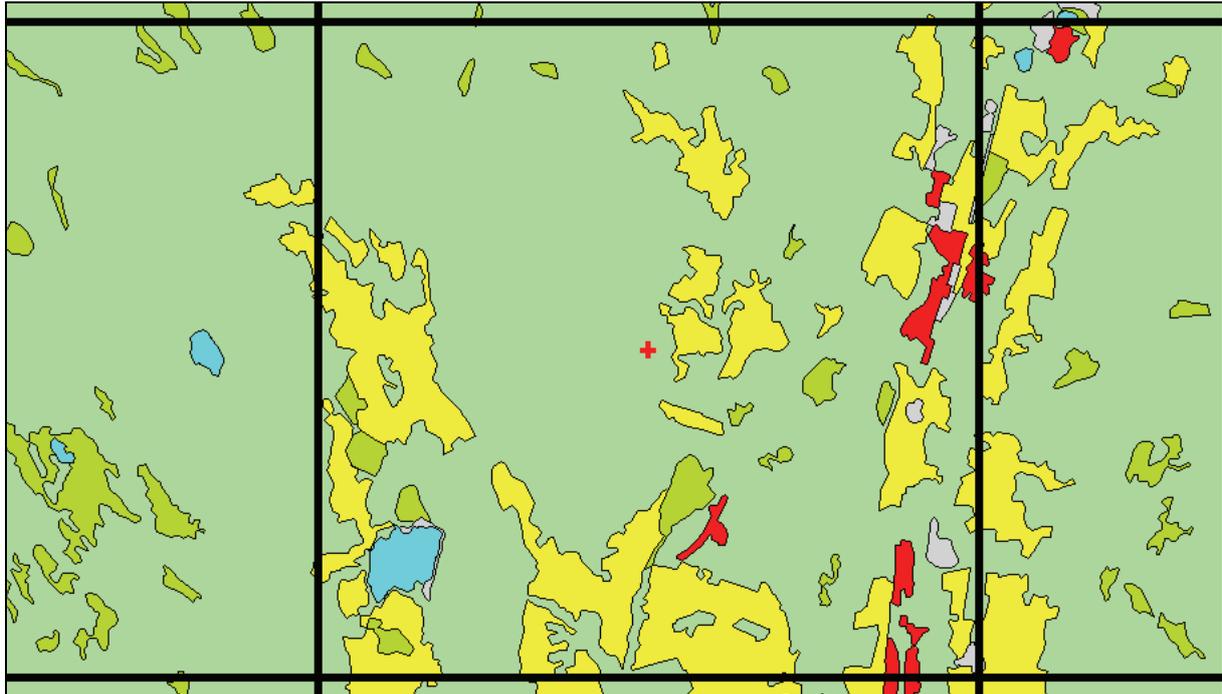


Figure 14 shows close up of one single grid cell in SSB5KM superimposed on the land resource map of Eidskog. In order to *measure* thematic classes the thematic polygons are clipped along the grid cell boundaries and the area of each class inside each grid cell is measured. This *overlay* is a GIS operation, but the resulting data table must also be processed in a tabular or statistical tool before the data can be attached to the grid.

Figure 15: Table with land resource types and area measures.

SSBID	Areatype	A10	A20	A30	A50	A60	A80	Total	F10	F20	F30	F50	F60	F80
23200006635000		0	0	2981463	0	543447	21270	3546180	0.00	0.00	84.08	0.00	15.32	0.60
23250006635000		0	0	430553	0	9409	0	439962	0.00	0.00	97.86	0.00	2.14	0.00
23200006640000	30	0	250467	12155139	0	1193510	914511	14513627	0.00	1.73	83.75	0.00	8.22	6.30
23250006640000	80	0	91152	11481090	84213	943849	2780465	15380769	0.00	0.59	74.65	0.55	6.14	18.08
23300006640000		0	308828	5862929	76498	129238	1735463	8112956	0.00	3.81	72.27	0.94	1.59	21.39
23350006640000		0	396268	8841961	0	521441	1583641	11343311	0.00	3.49	77.95	0.00	4.60	13.96
23400006640000		0	101217	6731284	0	107590	463992	7404083	0.00	1.37	90.91	0.00	1.45	6.27
23200006645000	30	0	127461	12941446	0	1072372	2040121	16181400	0.00	0.79	79.98	0.00	6.63	12.61
23250006645000	30	0	284867	18399812	0	1759193	4556132	25000004	0.00	1.14	73.60	0.00	7.04	18.22
23300006645000	80	187166	2140803	20680454	23911	319011	1648652	24999997	0.75	8.56	82.72	0.10	1.28	6.59
23350006645000	30	0	1362421	21409053	21869	1021329	1185329	25000001	0.00	5.45	85.64	0.09	4.09	4.74
23400006645000	20	829677	5676140	16995063	327274	445684	28810	24302648	3.41	23.36	69.93	1.35	1.83	0.12
23450006645000		49385	910600	7327694	43551	1439497	219188	9989915	0.49	9.12	73.35	0.44	14.41	2.19
23500006645000		0	0	120499	0	19009	0	139508	0.00	0.00	86.37	0.00	13.63	0.00
23200006650000	30	0	72373	15380924	0	567476	1314339	17335112	0.00	0.42	88.73	0.00	3.27	7.58
23250006650000	30	0	175420	22725003	0	1829392	270185	25000000	0.00	0.70	90.90	0.00	7.32	1.08
23300006650000	30	0	1235596	22681638	17231	625866	439667	24999996	0.00	4.94	90.73	0.07	2.50	1.76
23350006650000	30	796980	7125509	14288492	235089	1696131	857795	24999996	3.19	28.50	57.15	0.94	6.78	3.43
23400006650000	30	398992	7572272	14502273	248784	645196	1632483	25000000	1.60	30.29	58.01	1.00	2.58	6.53
23450006650000	30	0	1930750	20657636	77402	957339	1376871	24999996	0.00	7.72	82.63	0.31	3.83	5.51
23500006650000		21379	1127545	7533496	46902	304240	278942	9312504	0.23	12.11	80.90	0.50	3.27	3.00
23200006655000		0	29971	7273580	0	345287	900403	8549241	0.00	0.35	85.08	0.00	4.04	10.53
23250006655000	30	0	44954	22389475	15703	1845653	704211	24999996	0.00	0.18	89.56	0.06	7.38	2.82
23300006655000	30	0	153881	21301832	0	3162315	381972	25000000	0.00	0.62	85.21	0.00	12.65	1.53

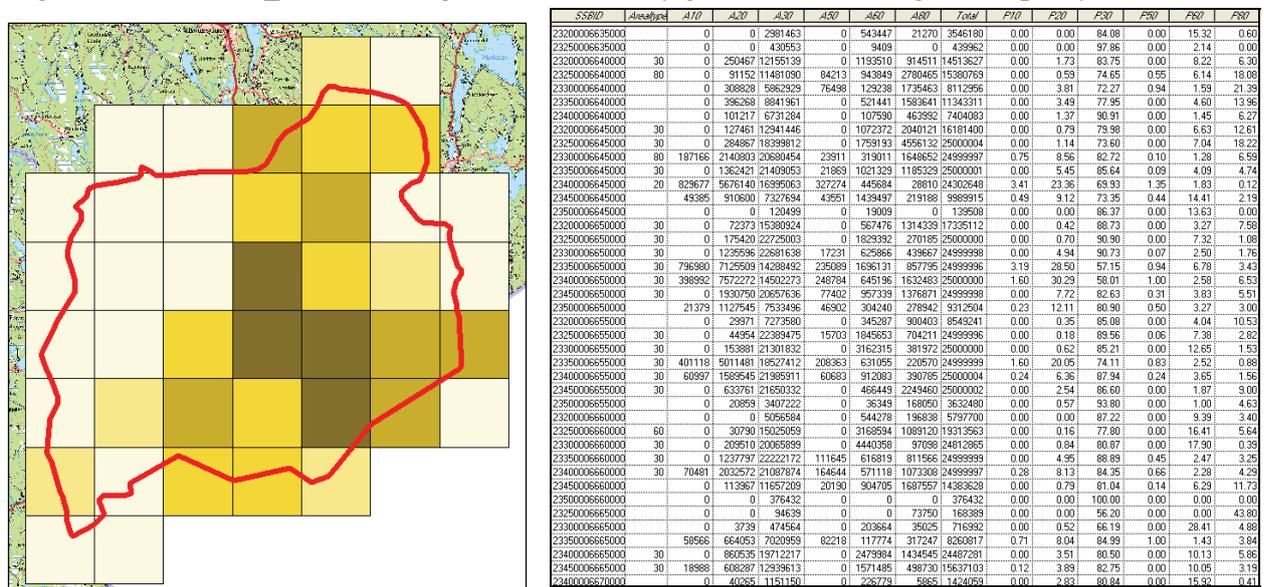
The basis for the area measurement method is a grid represented as a *polygon* dataset, i.e. each grid cell is represented as a polygon with the ID as identification. The grid polygon dataset is then joined with the new discrete polygon dataset (the *attribute* polygon dataset, e.g. land cover) in an overlay operation, some times known as a *spatial union*. The result is a new polygon dataset with small polygons each having a unique reference to a single grid cell and a single attribute polygon. Finally,

the area of each of these small polygons is calculated and the dataset is ready for transformation to an SSBgrid.

After tabular processing the result of the overlay can be attached to the grid cells (see figure 16). Here, each grid cell (SSBID) has a land resource type at the centre (Arealtype) and a total area (measured in m²) of various land resource types (A10 through A80). Total is the total area of the grid cell located within the municipal boundary. P10 through P80 are the area of each land resource class as a percentage of the grid cell.

In the transformation to an SSBgrid, a new attribute is calculated for each class of interest in the attribute polygon data set (e.g. for each land cover class). The values of these new attributes are initially set to 0. Then, for each of the small polygons, the area value is copied into the new attribute corresponding to its attribute class. Finally, the dataset is aggregated according to the method described in section 4.2 above, with the SSBID as the break variable and with “sum” as the descriptive function carried out on each of the new attributes.

Figure 16: SSB5KM_K0420 and agricultural area by grid cell in Eidskog municipality. Per cent.



Note: SSB5KM showing the extent of agricultural land within each cell, ranging from light yellow (less agricultural land) to dark brown (more agricultural land). Unlike figure 13 every cell has a single point within Eidskog municipality is counted. The share of agricultural land in cells along the borderline is only counted for the part of the cell located within Eidskog municipality. Base map: N250 Raster © Norge digitalt

Figure 16 shows SSB5KM with the percentage agricultural area in each grid cell in Eidskog (Attribute P20). Darker colour shows higher percentage. The grid cells are in this case much too large to convey a meaningful map and the example is only included to illustrate the method.

The area measurement method was used to compute the agricultural area for the grid cells in figure 16 above. A spatial union between SSB1KM and the Norwegian land cover map in scale 1:5,000 (Digitalt markslag) gave the basis for calculating the area of agricultural land within each SSB1KM grid cell.

4.5 Spatial filters

Spatial filters are neighbourhood operations carried out on existing grids. They require that the grid is organized as a matrix (a raster) and are easily applied using image processing software. In the filter

operation, each grid cell is assigned a value computed from the values found in the surrounding grid cells.

Figure 3 is an example of a spatial filter operation. The figure is using SSB1KM_F01. First population figures were assigned to each grid cell. Then, for each cell, the total population of itself and its eight closest neighbour cells were added up. Finally the figure was divided by 9 (nine cells = 9 km² in SSB1KM) in order to find the population density of the grid cell and its surroundings. An introduction to spatial filters is found in Lo and Yeung (2002) as well as in most other introductory textbooks on GIS or Image analysis.

4.6 Interpolation

Interpolation is used when the data value is known for some grid cells or for single points in the study area, and has to be estimated for the remaining grid cells. The situation is common in meteorology where data is obtained at scattered locations. An introduction to interpolation is found in Lo and Yeung (2002) as well as in most other introductory textbooks on GIS or Image analysis.

4.7 Local statistics

Global statistics like correlation coefficients and regression parameters can hide substantial regional differences or spatial variation. Local statistics is an approach where the statistics are calculated locally, i.e. individually for each grid cell. Clearly, each incident of the statistics must be calculated using a subset of the total population of grid cells. In this respect, local statistics is an advanced version of the *spatial filters* described above. The subset consists of data from the grid cells within a predefined distance from the target grid cell.

An example of the use of local statistics would be to compute the correlation between income and education for each grid cell in an SSBgrid using data from the neighbouring grid cells. The correlation coefficient is added to each grid cell as a new attribute and will (usually) have a spatial distribution. A thematic map of the resulting SSBgrid attribute can be drawn to visualize the results and provide the basis for further interpretation and discussion of the results. The result would amount to a thematic map of the correlation between the two variables. An introduction to local statistics is found in Fotheringham et al (2000).

5. Data sources

The registers from which the grid statistics is extracted must be geo-referenced by coordinates. The following registers fulfil the requirements and have been used in producing grid-maps in Statistics Norway.

5.1 Population and attributes

The central register for the population (DSF) includes all inhabitants in Norway. The register does not include coordinates but can be linked to the official address register (GAB – see next chapter) by a numerical address identifier. Because of a lack of concurrence in numerical address between the registers, not all of the population is identified by coordinates. In addition a few addresses in GAB do not have coordinates or the coordinates are wrong. However; as of 1st of January 2008, 99.7 per cent of the population were identified by coordinates.

Several variables are linked to every person in Norway; i.e. age, sex, education, income, health etc., and gives the option of combining several sets of information attributed to grid cells.

5.2 Buildings and attributes

The official Register for Ground Properties, Addresses and Buildings (GAB) consists of three mutually linked registers where the A- and G-part comprise all addresses and ground-properties. The B-part comprises information of all buildings larger than 15 m² including their co-ordinates. The G-part comprises information about all ground-property size etc. The register is under the responsibility of the Norwegian Mapping Authority. Each address, building and ground property is located by a point coordinate. Buildings have information on building type, area of largest floor, and more.

5.3 Ground properties and attributes

In 2009 the national register and map for ground properties will be integrated to one system called Matrikkelen (eng.: Matricle). This will lead to a harmonised national register with information on both the societal (i.e. ownership, sales values, use of properties) and the natural side (delineation of properties) of ground properties. Matrikkelen will be a profound source for developing new statistics, where also grid statistics will be possible.

5.4 Enterprises and attributes

The Central Register of Establishments and Enterprises (CRE) is a statistical register administered by Statistics Norway. Each relevant enterprise is identified by coordinates. The CRE include all relevant enterprises and establishments in both the private and public sector. The register has information on: organisation number, name, address, code of industry (NACE), number of employees, turnover, status/activity and more. In the initial geo-referencing project 75.4 per cent of the enterprises were identified by coordinates (Myro and Torp, 2002). The overall geo-referencing has later been improved to about 80 per cent.

5.5 The administrative farms register

The administrative farms register comprise all ground properties of agriculture and forestry with at least 5 000 m² land for agriculture and / or at least 25 000 m² forestry land (regardless of activity). The administrative farms register acts as a basic unit register with key identification to other administrative

registers on activity and subsidy arrangements. The administrative farms register is geo-referenced by coordinates (central farm building or the geographically central point on the ground property).

5.6 Norge digitalt

All datasets in the national geospatial data infrastructure (Norge digitalt) are geo-referenced and can be linked to grids using the methods described in chapter 4 above. Among the datasets are the national road data base, land use maps in scale 1:5 000 (AR5), registers for fresh water, the national map series in 1:50 000, forest taxation data, and the sport facilities register.

6. Considerations

6.1 Confidentiality

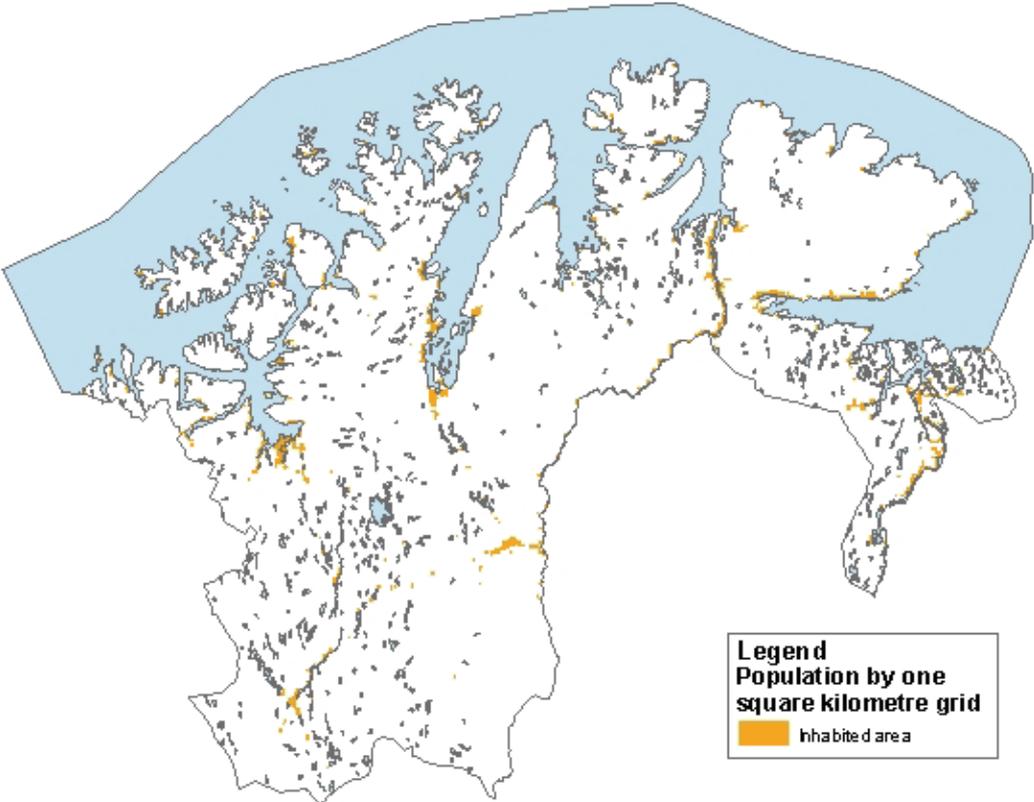
All datasets in the national geospatial data infrastructure (Norge digitalt) are geo-referenced and can be used in presentation of detailed statistics. The Statistical Act § 2-6 imply that Statistics Norway as a main rule do not publish tables with less than 3 units within a group (table cell) where the sample can lead to identification of individuals.

By for instance using population data on a 1x1 square km grid, it means that a lot of groups (here: grid cells) will have 1 or 2 observations. Choosing this grid cell size implies that there will be about 3 400 and 3 800 cells with 1 or 2 persons respectively. In total there are about 55 500 inhabited grid cells in Norway. The 7 200 cells with 1 or 2 persons constitutes 12.8 per cent of the inhabited grid cells. In comparison there are about 13 400 inhabited *local area units* (grunnkretser).

Cells with 1 or 2 observations are spread unevenly between urban and rural areas. A comparison of the sparsely populated county of Finnmark (the northernmost county), with the more densely populated county of Akershus, show that almost every fifth habited cell have less than 3 persons in Finnmark, while the share is 7 per cent in Akershus county.

When combining different geographical information for the unique grid cell, problems with confidentiality may increase further. The grid might for example be combined with for instance digital municipal borders or a digital road network which will give information on where the cells are located.

Figure 17. Population by 1 square kilometre grid. Whole of Finnmark.



If these grid cells are to be anonymous, there are several issues to be considered. Should one ensure that summing up values from the grid leads to the correct population figure for the country, or should one try to display all inhabited cells without necessarily giving the exact figure for every cell? Is it acceptable to use 3 observations in one grid cell as a universal lowest value whatever subject to illustrate, or should the number of observations be increased in order to ensure that it is anonymous?

The board of confidentiality has discussed several different solutions to this problem:

- Suppression method
- Heldals method
- Least value method
- Larger grid cells
- Clustering method
- Average method

Suppression of figures is a common method used at Statistics Norway. This method imply however that there will be a discrepancy from previous sums. If a grid cell contains the values 1 or 2, the value is set to be 0. Using population on 1 square kilometre grid as an example, it means that about 7 200 fewer grid cells will be inhabited.

Heldals method is used at Statistics Norway to make local area units with 1 or 2 observations anonymous. In short; cells with 1 or 2 observations are changed to 0 and 3, in such a way that the total number of observations is almost unchanged. Using this method on population data, about 3 500 grid cells will change from being inhabited to being uninhabited. Because each grid cell has its own unique ID, the method might be used the same way as with local area units.

Setting least values to 3, would for the population case increase the population figure with about 10 500 for the whole country. All inhabited grid cells would however be displayed.

All the three methods above satisfy the criteria that there should be at least 3 observations in a cell before data might be published; meaning that no individuals may be identified in the data set.

Increasing the grid cell size from 1x1 km to 2x2 km will reduce the number of cells with 1 or 2 observations quite much. Increasing grid cell size is however no solution in principle. The problem with confidentiality is practically spoken dependent on size of grid cells, but the principles with confidentiality can not be completely solved using this method.

In Denmark a method of *clustering grid cells* (joined) is used when confidentiality problems arises. This gives grid cells with differing sizes and shapes. The method is powerful, but implies a greater change in grid net and software.

Using average method means giving cells with value 1, 2 or 3 the average value for all cells containing these values. In this way all inhabited cells are displayed. Cells with value 3 are however changed to an average value. The average value might depend on regional level.

All these methods might be used, but with different results. The board of confidentiality at Statistics Norway were asked (1st of November 2006) to consider the different methods for handling confidentiality, and to set a criteria for disclosure control. The different methods mentioned above were discussed. Using methods which do not display all inhabited grid cells were considered to be a rather unsatisfactory solution for the users. Methods that change the grid size and/or shapes were also said to be less user friendly. Use of limits or cut-off values for individuals, is problematic with regard to households. If other variables than residents are to be used, one should consider to set a limit at number of inhabited addresses in each grid cell.

At the moment there is no satisfactory overview of households in Norway, and hence it is not suitable to operate with limit values for number of households per grid cell. An alternative is to set number of individuals within each cell so high that it normally would include more than 1 household.

Seen isolated it would not be in conflict with the confidentiality rules to say that there is only one resident within a square kilometre. However, if it on a later stage should be of interest to add other socio-economic variables, it would be harder to release this information when there already exists exact number of persons in a cell when the value is that low.

The board of confidentiality at Statistics Norway has only been asked to consider confidentiality with regard to population statistics (number of persons), without other variables. With respect to following cases, where need for additional variables will be brought up, it was suggested to set a high enough limit for number of persons with no other variables.

The board of confidentiality at Statistics Norway concluded that; for publishing of population statistics (number of persons) per 1 square kilometre, it should for the time being, not be given exact values for grid cells that contain less than 10 persons. It means that grid cells will operate with the values 0, 1-9, 10, 11 and 12 and so on. For grid cells with 1-9 persons the mean value is set to 5. To get up with a set of confidentiality rules for other themes, one has to gain more experience.

6.2 Non-equal areas

UTM33/EUREF89 is not an equal area projection and the size of a grid cell will vary slightly across the map, depending on the distance from the central meridian of the transverse Mercator projection (15° east for UTM-33). As an example, the true area of a grid cell in SSB1KM will be approximately 1.0008 km² along the 15° east meridian, 0.9917 km² in Kirkenes and 0.9926 km² in Utsira. This amounts to an error varying between 0.0 and 0.8 per cent caused by the relationship between the map projection and the curvature of the earth.

The effect of non-equal areas is not considered as a problem. Gridded data is in every respect a simplification of data from other and more detailed sources. Statistics requiring high accuracy should be compiled using the original sources, not the gridded representation. The most practical use of the gridded data will be for visualization and for national models combining data from multiple sources. The variable error in area estimate caused by the projection should be known by the user, but does not justify rejection of an otherwise pragmatic system. Compared with other sources of insecurity, the error is small, however systematic.

6.3 Compatibility with international grids

The grids defined in this document are intended for national use in Norway. Other grids will be required for international use. Examples on different grids in use are numerous. We hence only give a few examples. The Protocol on Long-term Financing of the Cooperative Program for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe (EMEP) is using two different grids (one with 50x50 km cells and one with 150x150 km cells) both based on a polar-stereographic projection (www.emep.int). The Joint Research Centre of the European Union and the European Environmental Agency are using a grid based on the Lambert Azimutal Projection (ETRS-LAEA) as described in EEA (2006). In order to deliver national data to these or similar international grids, transformation can be done on demand following a method approved by the receiving agency.

6.4 Metadata

Metadata is information about data. The more standardized the structure and content of information, the more effectively it can be used by both humans and machines. A metadata standard is simply a common set of terms and definitions that are presented in a structured format.

Standard of Spatial Information (SOSI) is the national standard for geo-referenced data in Norway. This standard is lead by a secretariat run under the Norwegian Mapping Agency, in cooperation with other institutions that deal with spatial information in Norway.

The international community, through the International Organization of Standards (ISO), has developed and approved an international metadata standard, ISO 19115. As a member of ISO, Norway is required to revise the SOSI-standard in accord with ISO 19115. Each nation can craft their own profile of ISO 19115 with the requirement that it include the 13 core element (Table 2).

Table 2: ISO 19115 Core Metadata Elements:

Mandatory Elements:	Conditional Elements:
Dataset title	Dataset responsible party
Dataset reference date	Geographic location by coordinates
Dataset language	Dataset character set
Dataset topic category	Spatial resolution
Abstract	Distribution format
Metadata point of contact	Spatial representation type
Metadata date stamp	Reference system
	Lineage statement
	On-line Resource
	Metadata file identifier
	Metadata standard name
	Metadata standard version
	Metadata language
	Metadata character set

Table 3: ISO 19115 Topic Categories

farming	geoscientificInformation	planningCadastre
biota	health	society
boundaries	imageryBaseMapsEarthCover	structure
climatologyMeteorologyAtmosphere	intelligenceMilitary	transportation
economy	inlandWaters	utilitiesCommunication
elevation	location	
environment	oceans	

Statistics Norway is working with standardising metadata for statistics attached to SSBgrid according to ISO 19115.

Table 4: Explanation of symbols

Category not applicable	.
Data not available	..
Data not yet available	...
Not for publication	:
Nil	-
Less than 0.5 of unit employed	0
Less than 0.05 of unit employed	0,0
Provisional or preliminary figure	*
Break in the homogeneity of vertical series	--
Break in the homogeneity of horizontal series	

In addition to the ISO Core Metadata Elements listed above, Dataset topic categories shall also be

provided in the metadataset. This set of 19 high-level subject categories (Table 3) is provided within the ISO 19115 standard as a means of quickly sorting and accessing thematic information.

To avoid double work, use of existing metadata and standards attached to existing statistics will be used in the further work. One of the standards in use is the use of symbols in tables (Table 4). Where suitable these symbols should also be used in grid statistics, where each grid cell may be considered as a cell in an ordinary statistical table.

7 References

- Bye, Anne Snellingen, Sandmo, Trond og Berge, Gisle (2006): Jordbruk og miljø. Resultatkontroll jordbruk 2006. Rapport 2006/ 37. Statistisk sentralbyrå 2006.
- EEA (2006): Guide to geographical data and maps, *EEA operational guidelines*, January 2006, version 2.0 European Environmental Agency, Copenhagen 2006
- Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2000) *Quantitative Geography. Perspectives on Spatial Data Analysis*. SAGE Publication, London 2000
- GridClub of Europe (2007). <http://www.gridclub.ssb.no/> , redirected to <http://www.efgs.ssb.no/>
- European Forum for GeoStatistics (2009) <http://www.efgs.ssb.no/>
- Heldal, Johan and Fosen, Johan. (2001): "Statistisk konfidensialitet i SSB. Et diskusjonsnotat", Notater 2001/28, Statistisk sentralbyrå.
- Lie, Einar and Roll-Hansen (2001). *Faktisk talt. Statistikkens historie i Norge*. ISBN: 9788215001500. Universitetsforlaget 2001.
- Lo, C.P. and Yeung, A.K.W. (2002) *Concepts and Techniques of Geographic Information Systems*, Prentice Hall, New Jersey 2002
- Myro, Svein and Torp, Camilla (2002): Stedfesting av bedrifter i Bedrifts- og foretaksregisteret. Hovedprosjekt. Notater 2002/ 27. Statistisk sentralbyrå 2002.
- Nordic Forum for Geo- statistics (2006). Norway, Kongsvinger, Skansgården 12th - 14th September 2006. http://www.ssb.no/omssb/kurs_seminar/nordic/
- Ottestad, Arne Knut (2006). *Metode for presentasjon av statistikk på rutenett. (Method for presentation of statistics on grid)*. Note to Board of confidentiality at Statistics Norway. 22.09.2006.
- Statistics Finland (1995): "Differences in the spatial structure of the population between Finland and Sweden in 1995". Research report 221.
- St.mld. nr. 30: 2002-2003. "Norge digitalt" – Et felles fundament for verdiskaping.
- Strand, Geir-Harald. and Rekdal, Yngve. (2006). Area frame survey of land resources - AR18×18 system description. Report 2006/03. The Norwegian Forest and Landscape Institute.
- Strand, G-H. and Rogstad, L. (2004). Befolkningstetthet i og omkring jordbruksarealet, *Kart og Plan* 64: 75-77
- Takle, Mona (2002): Befolkningsstatistikk på rutenett. Dokumentasjon. Notater 2002/ 22. Statistics Norway 2002. (Population statistics on grids.)
- Valen, Thorleiv (2006). *Minutes from meeting in Board of confidentiality at Statistics Norway*. 20.10.2006.

Appendix

Statistics by subject and ISO 19115 Topic Categories

About the statistics provides central information about every statistics published by Statistics Norway. This metadata information is sorted by subject. This table shows statistics sorted by subject and suggested coding of ISO 19115 Topic Categories. Note that some statistics may fall within several categories. This table suggests the main category of each statistics.

Statistics sorted by subject

ISO 19115 Topic Categories

00 General

00.00.10 National	society
00.00.20 Regional	society
00.01.10 General elections	society
00.01.20 Local elections	society
00.02 Living conditions	society
00.02.10 Gender equality	society
00.02.20 Time use	society

01 Natural resources and the environment

01.01.10 Land cover	environment
01.01.20 Land use	environment
01.03.10 Energy	society
01.04 Pollutions	environment
01.04.10 Air	environment
01.04.20 Water	society
01.05 Waste	society
01.05.10 Municipal waste	society
01.05.20 Industrial waste	society
01.05.30 Hazardous waste	society
01.05.40 Packaging waste	society
01.06.20 Environmental protection expenditures	economy

02 Population

02.01 Population structure	society
02.01.10 Population	society
02.01.20 Families, households	society
02.02 Vital statistics and migration	society
02.02.10 Births, adoptions, deaths, mortality	society
02.02.20 Migration, immigration and emigration	society
02.02.30 Marriages, divorces, and other relationships	society
02.03 Population projections	society

03 Health, social, welfare and crime

03.01 Health conditions	health
03.02 Health services	health
03.03.10 Child welfare	society
03.04 Social services, social security and social assistance	society
03.04.20 Social services outside institutions	society
03.04.30 Social security, social assistance and crime	society
03.05 Crime and the justice	society

04 Education

04.01 Level of education	society
04.02 Educational institutions	society
04.02.10 Day-care centres, preschools	society
04.02.20 Primary and lower secondary schools	society
04.02.30 Upper secondary schools	society
04.02.40 Universities and colleges	society
04.02.50 Adult education and other tuition	society

05 Personal economy and housing

05.01 Income, property, taxes	economy
05.02 Consumption	economy
05.03 Dwelling and housing conditions	society

06 Labour market, wages

06.01 Labour force participation	economy
06.02 Working conditions, sickness absenteeism	society
06.03 Unemployment	economy
06.04 Labour disputes	society
06.05 Wages and labour costs	economy

07 Culture

07.01.10 Radio and television	society
07.01.40 Libraries, museums, archives, and other dissemination of culture	society
07.02.10 Membership in organizations, religious and philosophical communities	society
07.02.30 Media use, book reading, library visits	society
07.02.50 Sport and outdoor recreation	society

08 Price indices and economic indicators

08.02 Price indices	economy
08.02.10 Consumer price index	economy
08.02.20 Manufacturer and wholesale price indices	economy
08.02.30 House price indices and construction cost indices	economy
08.03.10 New orders, reserves of unfilled orders	economy
08.03.20 Turnover	economy
08.03.30 Investment	economy
08.03.40 External trade	economy
08.04 Production indices	economy
08.05.10 National	economy

09 National accounts and external trade

09.01 National accounts	economy
09.02 Financial sector balances	economy
09.03 Balance of payments	economy
09.04 Foreign assets and liabilities, foreign ownership interests	economy
09.05 External trade	economy

10 Industrial activities

10 Industrial activities	economy
10.01 Structure, business register	economy
10.02 Business taxes	economy
10.03 Technological indicators, incl. ICT	economy
10.04 Agriculture and forestry	economy
10.04.10 Agriculture, hunting, wildlife conservation	economy
10.04.20 Forestry	economy
10.05 Fishing and fish farming	economy
10.06.20 Extraction of crude oil and natural gas	economy
10.07 Manufacturing	economy
10.08.10 Electricity, gas, steam and hot water supply	economy
10.09 Construction	economy
10.10 Wholesale and retail trade, repair activities	economy

10.10.10 Wholesale and retail trade	economy
10.11 Hotels and restaurants	economy
10.12 Transport and communication	economy
10.12.10 Rail transport	economy
10.12.20 Road transport	economy
10.12.40 Sea transport	economy
10.12.50 Air transport	economy
10.12.60 Transport and travel agency-related services	economy
10.12.70 Post and telecommunications	economy
10.13.10 Financial institutions	economy
10.13.20 Other Financial enterprises	economy
10.14 Real estate, renting and business activities	economy
10.14.10 Sale and management of real estate	economy
10.16 Other service activities	economy

11 Financial markets

11.01 Securities, credit indicators and monetary aggregates	economy
11.02 Bankruptcies	economy

12 Public finances

12.01 Public finances and social security schemes	economy
12.01.10 State finances	economy
12.01.20 Municipal finances	economy
12.01.30 Tax accounts	economy